
Volume 32, Issue 1, January/February 2011 ISSN 0891-4222
ELSEVIER
Research in Developmental Disabilities

Research in Developmental Disabilities 32 (2011) 180–187

Contents lists available at ScienceDirect

Research in Developmental Disabilities

How much change is true change? The smallest detectable difference of the Preschool Imitation and Praxis Scale (PIPS) in preschoolers with intellectual disabilities of heterogeneous aetiology

M. Vanvuchelen [a,b,c,*], C. Vochten [d]

[a] Department Health Care, PHL University College, Belgium
[b] Department of Rehabilitation Sciences, V.U.B., Belgium
[c] Department of Rehabilitation Sciences, K.U.Leuven, Belgium
[d] Student Counseling and Guidance Center, Hasselt, Belgium

ARTICLE INFO

Article history:
Received 11 September 2010
Accepted 20 September 2010

Keywords:
Imitation
Preschool child
Intellectual disabilities
Reliability
Smallest detectable difference
Preschool Imitation and Praxis Scale (PIPS)

ABSTRACT

The teaching of imitation skills is often the first step in interventions for young learners with intellectual disabilities. The main goal of this study was to determine the smallest detectable difference (SDD) at 95% confidence of the Preschool Imitation and Praxis Scale (PIPS) in preschoolers with intellectual disabilities. Two raters independently scored videotapes of the imitation performance of 44 preschoolers (27 with Down syndrome, 10 with Non-Specific Mental Retardation and 7 with Low-functioning Autism) between 13 and 58 months of age (mean age 39.6 months, SD 11.9 months). Results revealed that the PIPS demonstrated acceptable interrater reliability on item level (weighted kappa values ranged from 0.52 to 0.96) and scale level (ICC = 0.986; 95% CI: 0.975–0.993). The SDD of the PIPS was 7.2%, indicating that the change score rated by different raters for an individual child with an intellectual disability is valid and that the PIPS can be used by early interventionists and researchers as an outcome measure to determine children's maturation or improvement.

1. Introduction

Motor imitation (hereafter 'imitation') is a powerful mechanism to learn motor skills and to facilitate comprehension of other persons' actions and mental states (Hurley & Chater, 2005). The term imitation has many definitions (Sevlever & Gillis, 2010). There are however two connotations in which it is most commonly used. The first defines imitation as the capacity of an individual to replicate an observed motor act (Prinz, 2002). The second defines imitation as the capacity to acquire, by observation, a new motor behaviour and to repeat it using the same movements employed by the demonstrator (Tomasello, Carpenter, Call, Behne, & Moll, 2005).

Imitation skills occur without direct training in typically developing children and play an important part in the development of age-appropriate praxis and social skills (Masur, 2006). Imitation avoids time-consuming trial-and-error learning. In reproducing the exact and detailed features of the demonstrator's actions, children are likely to successfully complete the intended actions, even with a limited understanding of their purpose. Faithful copying can be used to disentangle the goal of an action to be imitated when it is not completely clear to the child (Lyons, Young, & Keil, 2007).

* Corresponding author at: Sterrebos 111, B 3512 Stevoort, Belgium. Tel.: +32 11274124; mobile: +32 497449940.
E-mail addresses: MVanvuchelen@mail.phl.be, Marleen.Vanvuchelen@faber.kuleuven.be (M. Vanvuchelen).

0891-4222/$ – see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.ridd.2010.09.019

Consequently, the precision of the perception-action coupling, expressed in the accuracy of imitation performances is an important criterion for children's aptitude to learn new skills.

The acquisition of skilled acts, learned effortlessly by more able young children, is a crucial aspect of the overall development of children with developmental disabilities. The imitation problems among learners with intellectual disabilities (Macedoni-Luksic et al., 2009; Vanvuchelen, Feys, & De Weerdt, in press) and with autism spectrum disorders (Perra et al., 2008; Vanvuchelen, Roeyers, & De Weerdt, in press-a; Vanvuchelen, Roeyers & De Weerdt, 2011a; Williams, Whiten, & Singh, 2004; Zachor, Ilanit, & Itzchak, 2010) may prevent them from learning the skills that typically developing children learn in natural environments. Adults cannot directly teach children all the skills needed for adult life. Appropriate learning depends primarily on the child himself or herself taking on the responsibility for learning throughout the day by imitating others, engaging others, and exploring the potential of the various environments. Therefore, the teaching of imitation skills is often the first step in interventions for young learners with intellectual disabilities (Baer, Peterson, & Sherman, 1967; Brown, Brown, & Poulson, 2008; Brown, Peace, & Parsons, 2009; Mehta, Pande, & Bhargava, 1991) and autism spectrum disorders (Lovaas, 1987; Stahmer, Ingersoll, & Carter, 2003; Rogers, 1998). Behavioural intervention approaches to teach imitation skills with the use of video modeling (Tereshko, MacDonald, & Ahearn, 2010) and live modeling, e.g., Discrete Trial Training (Brown et al., 2008) and Reciprocal Imitation Training (Cardon & Wilcox, in press; Ingersoll, 2010; Ingersoll & Gergans, 2007; Ingersoll & Lalonde, 2010; Ingersoll, Lewis, & Kroman, 2007; Ingersoll & Schreibman, 2006) are essential methods to improve the learning potential of intellectual disabled children with imitation problems. Although the existing literature on effects of imitation teaching provides evidence that this kind of early intervention strategy improves the prognosis of the children, research findings are based on methodologically weak studies with few participants and the absence of a control group, an alternative intervention strategy and a systematic imitation assessment.

The Preschool Imitation and Praxis Scale (PIPS) is designed to assess the accuracy of bodily (actions without objects) and procedural (actions with objects) imitation performance at preschool age (Vanvuchelen, Roeyers, & De Weerdt, 2011b). Imitation age-equivalent scores are derived from PIPS scores of 654 typically developing children between 12 and 59 months of age (Vanvuchelen, 2009). Evidence of the validity and reliability of the PIPS in healthy preschoolers has been provided in several studies (Vanvuchelen et al., 2011b; Vanvuchelen, Roeyers, & De Weerdt, in press-b). However, reliability that is established on individuals from one population, e.g., typically developing preschoolers, cannot automatically be attributed to another population, e.g., preschoolers with intellectual disabilities. This has been termed population-specific reliability. The factor of intellectual disability can alter the way a child responds to a measurement and the consistency with which examiners can take the measurement (Portney & Watkins, 2009).

Furthermore, to be clinically meaningful an imitation assessment must be reliable enough to evaluate outcomes of a therapeutic intervention, such as teaching imitation skills. Knowing that the PIPS has achieved a certain level of psychometric adequacy to identify preschoolers with imitation problems (Vanvuchelen et al., in press-a) says nothing about its sensitivity to treatment-related changes in child functioning. The intent of the assessment, e.g., making a diagnosis versus testing the efficacy of an intervention, can potentially result in a different estimate of important change (Norman, Sloan, & Wyrwich, 2003). To determine if the imitation aptitude has changed, an examiner must know what part of the difference between children's measurements is attributable to real change, and what part is due to measurement error. Reliability can be reported in relative or absolute terms. Relative interrater reliability statistics indicate the degree of association between two or more ratings (e.g., intraclass correlation coefficients or ICCs), but they do not provide clinical guidance for assessing real changes at an individual child level. If the PIPS is used in intervention studies of preschoolers with intellectual disabilities, the knowledge of the standard error of measurement (SEM) and the smallest detectable difference (SDD) is important to ascertain which children show a 'true' imitation change apart from differences due to measurement errors (Atkinson & Nevill, 1998; Portney & Watkins, 2009). The SEM and SDD provide indications of absolute reliability. The precision of the measurement is expressed as the SEM. The SDD is the amount of change in a variable that must be achieved to reflect a true difference (Portney & Watkins, 2009). This is the smallest amount of difference that passes the threshold of error for a specific instrument, e.g., the PIPS, and application, e.g., in preschoolers with intellectual disabilities. The present article provides evidence of relative and absolute interrater reliability of the PIPS in preschoolers with intellectual disabilities of heterogeneous aetiology.

2. Methods

2.1. Participants

Forty-four preschoolers (19 female and 25 male) between 13 and 58 months of age (mean chronological age 39.6 m, SD 11.9 m) with an intellectual disability (27 with Down syndrome, 10 with Non-Specific Mental Retardation and 7 with Low-functioning Autism) participated in this study. Nonverbal mental age was determined with the Dutch modification of the Bayley Scales of Infant Development (BSID-II-NL; Van der Meulen, Ruiter, Lutje Spelberg, & Smrkovský, 2000) and ranged between 8 and 43 months of age (mean nonverbal mental age 25.6 m, SD 9.7 m).

The children with Down syndrome and with Non-Specific Mental Retardation were recruited from home-based child development programs for children with intellectual disabilities, respectively special schools for children with learning disabilities. The children with Low-functioning Autism were diagnosed according to a multidisciplinary clinical consensus classification (University Autism Clinics) in addition to a positive ADOS-G-classification (Lord, Rutter, DiLavore, & Risi, 2003). All families gave written informed consent for the participation of their child.

2.2. Preschool Imitation and Praxis Scale (PIPS)

The theoretical framework and deductive test construction approach of the Preschool Imitation and Praxis Scale (PIPS) have been described in detail elsewhere (Vanvuchelen et al., 2011b). To recap briefly, different action types considered to be important as unravelled in research in apraxic adults were selected to tap a broad range of possible imitation mechanisms: action types with different effects (salient environmental and internal), representational levels (meaningful and non-meaningful; goal directed and non-goal directed), temporal complexities (single and sequential) and visual monitoring possibilities (transparent and opaque). The PIPS consists of 10 task categories with each three tasks: six gestural, three procedural and one facial. Imitation tasks which are possible to be performed by young children but unlikely to be exhibited spontaneously were selected (Vanvuchelen et al., 2011b). The 10 task categories and 30 PIPS tasks are described in Table 1.

Imitation performances on each task are scored on a 3–5 point scale in accordance with the criteria of the scoring system of the PIPS, which evaluates the spatiotemporal resemblance between the modelled and copied action (Vanvuchelen, 2009). To illustrate this system, we exemplify the scoring of the task "to pretend to comb your hair with an imaginary comb". Score 4 is given if the child has used a symbolic grip and has performed a repetitive action on both sides of the head. Score 3 is given if the child has used a symbolic grip and has performed a repetitive action on one side of the head or a single action. Score 2 is given if the child has used a body-part-as-an-object grip and has performed a repetitive action on both sides of the head. Score 1 is given if the child has used a body-part-as-an-object grip and has performed a repetitive action on one side of the head or a single action. Score 0 is given is the child has performed another action or has refused to imitate. The final PIPS score (ranging between zero and 81) is a reflection of the accuracy of the child's imitation performance (Vanvuchelen, 2009). Interrater reliability of the PIPS items (weighted kappa values ranged from 0.64 to 1) and the total score (intraclass correlation coefficient ICC = 0.995; 95% CI: 0.990–0.997) has been established in a population of typically developing children. In this population, the SDD was 5.6% of the possible score range of the PIPS (Vanvuchelen et al., in press-b).

For the test administration, each child was individually assessed in a quiet room and was seated at a table in front of the trained examiner. The administration of the 30 tasks of the PIPS was in accordance with the guidelines for item instruction of the PIPS (Vanvuchelen, 2009). Before demonstrating each task, a child's attention was attracted by addressing the child by name. Only the verbal instruction "(Name), you do this too" was given.

2.3. Data analysis

All analyses have been performed using the statistical software SPSS (version 16.0). *P*-values smaller than 0.05 were considered as significant.

2.4. Descriptive statistics

Simple descriptive statistics (mean and standard deviation) are used to describe the basic features of the PIPS in this study.

2.5. Relative interrater reliability

To determine relative interrater reliability, the imitation performance of the children was videotaped. Two trained examiners scored the performances on the videotapes independently.

Relative interrater reliability on individual item level were examined using Cohen's weighted kappa values. The interpretation of kappa values was done according to Fleiss (Fleiss, 1981): below 0.40 agreement by coincidence; between 0.40 and 0.60 moderate; between 0.61 and 0.75 good and above 0.75 excellent agreement. Percentages of agreement were also determined for all items: 70% or higher was considered as acceptable.

Relative interrater reliability on scale level was examined with the intraclass correlation coefficient (ICC). The option 'treating subjects as well as raters as a random effect' was applied. This resulted in a two-way random effects model with measures of absolute agreement. The interpretation of the ICC was done according to Portney and Watkins (2009): a value below 0.75 is indicative of poor to moderate reliability and a value above 0.75 of good reliability. Bland and Altman's limits of agreement on the scale score (mean of the differences between ratings $\pm$2SD) were used to assess the strength of agreement between the raters (Bland & Altman, 1986).

2.6. Absolute interrater reliability

Absolute interrater reliability was examined by the smallest detectable difference (SDD) to determine whether the change scores rated by different raters of an individual child were valid (beyond random errors) at the 95% confidence level.

Table 1
Description of the 10 task categories and 30 items of the Preschool Imitation and Praxis Scale (PIPS) presented in standardised order.

Item nr	T	Category	Dimension	Task description	Score
1	P	sao-P1	GDP	Raise a toy bear by pulling a cord	0–2
2	P	sao-P2	GDP	Put a wooden block on top of your head	0–2
3	P	sao-P3	GDP	Switch on a lamp in a toy animal with your forehead	0–2
4	G	i-MG1	SIB	Perform the gesture to "wave good-bye"	0–2
5	G	i-MG2	SIB	Perform the gesture to "show something with an outstretched hand in supination"	0–2
6	G	i-MG3	SIB	Perform the gesture to "beckon with the index finger"	0–2
7	G	si-NMG1	SIB	Raise your outstretched arm till 90° anteflexion and make a circle with the index finger and thumb	0–3
8	G	si-NMG2	SIB	Raise your outstretched arm till 90° anteflexion and stretch out your fingers	0–3
9	G	si-NMG3	SIB	Raise your outstretched arm till 90° anteflexion, hold up the little finger while all the other fingers and the thumb are bent	0–3
10	P	sq-P1	NGDP	Open the box, put the lid on the table, turn the box upside-down, put the block on the bottom of the box	0–3
11	P	sq-P2	NGDP	Take the block from the bottom of the box, turn the box in normal position again, close the box, put the block on the lid of the box	0–3
12	P	sq-P3	NGDP	Take the block from the lid of the box, open the box, put a disc into the box, close the box, put the block again on the lid of the box	0–3
13	G	t-MG1	SIB	Pretend to "comb your hair with an imaginary comb"	0–4
14	G	t-MG2	SIB	Pretend to "open an imaginary door with an imaginary key"	0–4
15	G	t-MG3	SIB	Pretend to "brush your teeth with an imaginary toothbrush"	0–4
16	G	bi-NMG1	SIB	Place one fist on top of the other	0–3
17	G	bi-NMG2	SIB	Extend the index fingers of both hands while the other fingers and thumbs are bent, and bring the top of the index fingers towards each other	0–3
18	G	bi-NMG3	SIB	Open one hand in vertical position and touch the top of the fingers with the palm of the other hand in horizontal position	0–3
19	G	fa-NMG1	SIB	Extend your index finger and touch the top of your nose	0–3
20	G	fa-NMG2	SIB	Touch your lower lips with the nails of your thumbs	0–3
21	G	fa-NMG3	SIB	Extend the index finger of your left hand and touch your right cheek and extend the index finger of your right hand and touch your left cheek	0–3
22	P	aso-P1	GDP	Turn a cup upside-down and play drums on it with two spoons	0–2
23	P	aso-P2	GDP	Remove the cap of a doll and put a shoe on the head of the doll	0–2
24	P	aso-P3	GDP	Put a toy car in bed, turn it upside-down and tuck it in with a blanket	0–2
25	G	sq-NMG1	SQB	Hit the table with the palm of your hands, cross the arms and hit the table again, return to the original position and hit the table once more	0–3
26	G	sq-NMG2	SQB	Hit the table with one hand in supination, turn the hand in pronation and hit the table again, clap in the hands, hit the table with the palm of both hands	0–3
27	G	sq-NMG3	SQB	Hit the table with both hands in supination, turn the hands in pronation, hit the table again, clap in the hands, hit the table with the palm of both hands once more	0–3
28	F	f1	SIB	Shake the head, eyes closed to say 'no', with an expression of disapproval	0–2
29	F	f2	SIB	Look angry with a frown of the eyebrows	0–2
30	F	f3	SIB	Nod quickly with your head and show an expression of happiness	0–2

T, type of action (P, procedural imitation; G, gestural imitation; F, facial imitation); sao-P, substituted-actions-upon-objects; i-MG, intransitive meaningful gestures; si-NMG, single non-meaningful hand postures; sq-P, action-sequences-upon-objects; t-MG, transitive meaningful gestures; bi-NMG, bimanual non-meaningful hand postures; fa-NMG, non-meaningful hand postures to the face or head; aso-P, actions-upon-substituted-objects; sq-NMG, sequences of non-meaningful hand postures; f, facial expressions; GDP; goal directed procedural imitation; NGDP, non-goal directed procedural imitation; SIB, single bodily imitation; SQB, sequential bodily imitation.

First, the standard error of measurement (SEM) for the two separate ratings (SEM_{first} and SEM_{second}) was calculated on the basis of the standard deviation (SD) and the ICC: $SEM = SD \times \sqrt{1 - ICC}$, which was used to quantify the amount of random measurement errors. Assuming that measurement errors are distributed normally, the corresponding 95% confidence interval, in which the true score is expected, was $\pm 1.96 \times SEM$ (Portney & Watkins, 2009). There are no clear criteria for an acceptable SEM-value available. So, how much ratings can be deviant from one another will be a question of judgment. Preferably, this should be defined in advance (Bland & Altman, 1986). We expressed the SEM as a percentage of the possible score range of the PIPS (from 0 to 81). We postulated that a value lower than 10 percent of the possible score range of the PIPS, which means a SEM score below 8.1, and a difference between the percentages of two SEMs smaller than 1 percent would be satisfactory (Van Baalen, Odding, van Woensel, & Roebroeck, 2006).

Then, the smallest detectable difference (SDD) was calculated using the following formula: $SDD = 1.96 \times \sqrt{(SEM_{first}^2 + SEM_{second}^2)}$. If the difference between both SEMs was smaller than 1 percent, we assumed that they were equal and simplified the formula to: $SDD = 1.96 \times SEM \times \sqrt{2}$ (Portney & Watkins, 2009). When interpreting the SDD the following criteria were used. First, when the SDD value was lower than 10% of the possible highest score of the PIPS, which means a SDD score below 8.1, the PIPS was considered acceptable (Kropmans, Dijkstra, Stegenga, Stewart, & de Bont, 1999). Second, the SDD should be close to 0.5 standard deviation for discriminating the threshold of change (Norman et al., 2003).

3. Results

3.1. Descriptive statistics

The mean total PIPS score was 28.7 (SD 17.9) according to the first rater and 29.2 (SD 19.2) according to the second rater.

3.2. Relative interrater reliability

Table 2 provides the weighted kappa values and percentages of agreement for the interrater agreement of the 30 individual item scores. Among the 30 PIPS items, 18 revealed an excellent interrater agreement for individual item scores, 4 had good agreement, and 3 had moderate agreement. The kappa statistics of 5 items could not be computed, because of the skewed distribution of the data. However, the percentage of agreement of these 5 items was obtained: 70% (t-MG1), 72% (t-MG2), 70% (t-MG3), 84% (fa-NMG2) and 79% (sq-NMG1).

The interrater reliability of the PIPS scale scores was high (ICC = 0.986; 95% CI: 0.975–0.993). Fig. 1 plots limits of agreement calculated from interrater data of the PIPS scale scores. Plotting PIPS means of the two observations against the interrater difference for each participant does not give any indication that measurement errors vary systematically over the range of possible scores.

3.3. Absolute interrater reliability

The SEM of the first rater was 2.1 (2.6% of the possible score range of the PIPS) and of the second rater 2.3 (2.9% of the possible score range of the PIPS). According to the data of the first rater, there is a 95% chance that the true mean PIPS score lies between 24.5 and 32.2; and according to the second rater between 24.6 and 33.8. The difference between both was 0.2 (0.3% of the possible score range of the PIPS). Since the difference between the percentages of two SEMs was smaller than 1%, we have used the SEM of the first rater to determine the SDD. The SDD was 5.8 (7.2% of the possible score range of the PIPS).

Table 2
Weighted kappa values and percentages of agreement for interrater agreement of 30 individual item scores of the Preschool Imitation and Praxis Scale (PIPS).

Item number	Task	Weighted kappa value	Asymptotic standard error	Percentages of agreement
1	sao-P1	0.96	0.03	97.7%
2	sao-P2	0.78	0.08	86.4%
3	sao-P3	0.91	0.04	93.2%
4	i-MG1	0.81	0.07	86.4%
5	i-MG2	0.79	0.06	81.8%
6	i-MG3	0.85	0.05	86.4%
7	si-NMG1	0.77	0.08	86.4%
8	si-NMG2	0.71	0.09	84.1%
9	si-NMG3	0.83	0.06	86.4%
10	sq-P1	0.85	0.05	88.6%
11	sq-P2	0.85	0.05	88.6%
12	sq-P3	0.90	0.04	90.9%
13	t-MG1	[a]	[a]	70.4%
14	t-MG2	[a]	[a]	72.7%
15	t-MG3	[a]	[a]	70.4%
16	bi-NMG1	0.73	0.07	77.3%
17	bi-NMG2	0.68	0.08	77.3%
18	bi-NMG3	0.58	0.09	70.4%
19	fa-NMG1	0.58	0.09	70.4%
20	fa-NMG2	[a]	[a]	84.1%
21	fa-NMG3	0.86	0.07	93.2%
22	aso-P1	0.81	0.06	86.4%
23	aso-P2	0.92	0.05	95.4%
24	aso-P3	0.82	0.07	88.6%
25	sq-NMG1	[a]	[a]	79.5%
26	sq-NMG2	0.86	0.07	93.2%
27	sq-NMG3	0.73	0.09	84.1%
28	f1	0.79	0.06	84.1%
29	f2	0.52	0.10	72.7%
30	f3	0.84	0.06	88.6%

[a] Kappa statistics could not be computed, because of the skewed distribution of the data.

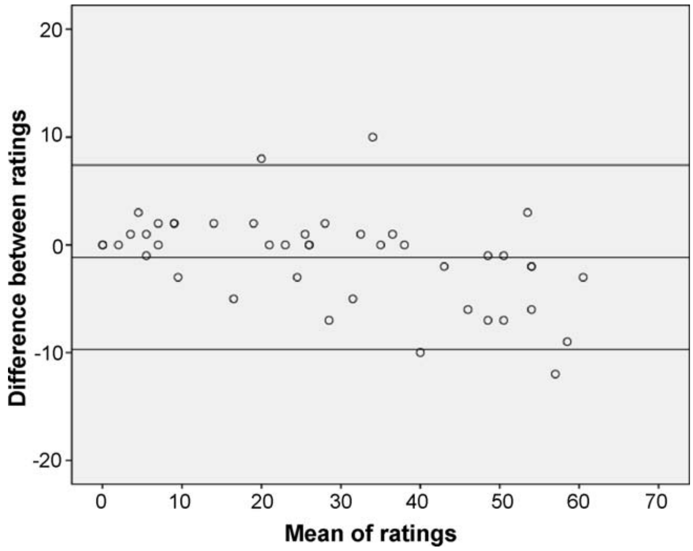

Fig. 1. Bland and Altman's limits of agreement for the Preschool Imitation and Praxis Scale (PIPS) scale score on the interrater data.

4. Discussion

Imitation is one of the pivotal skills that young children use to learn from others. Targeting the development of imitative skills in learners with imitation problems may be an important focus in early intervention. Interventions such as teaching imitation skills that result in child initiation, exploration, and ongoing engagement of the social as well as the physical world are likely to lead to greater child learning long term. These self-learning behaviours need to be identified as treatment goals and examined when assessing children's response to treatment (Rogers, 1998).

The Preschool Imitation and Praxis Scale (PIPS) is a reliable and valid measure to identify preschoolers with imitation problems (Vanvuchelen, Feys et al., in press; Vanvuchelen et al., in press-a). To be clinically meaningful the PIPS must be reliable enough to evaluate outcomes of an intervention, such as teaching imitation skills. In clinical settings, it is essential to know by how much the score of one rater is likely to differ from the score of another rater. If this is not enough to cause problems in clinical interpretation we can use the two interchangeably. The present paper provides evidence for population-specific reliability, in particularly in preschoolers with intellectual disabilities of heterogeneous aetiology. Random errors reduce the objectivity of the data and can arise from inconsistencies in the administering and scoring protocol. Therefore, the examiners were trained in these protocols to prevent inconsistencies of scores. The interrater reliability was moderate to excellent at the individual item level and good for the PIPS total score.

However, the interrater reliability based on correlation coefficients, such as an intraclass correlation coefficient, provide an indication of relative reliability (Atkinson & Nevill, 1998). A high intraclass correlation does not always mean that the two raters agree. Correlations measure the strength of a relation between two scores, not the agreement between them. When data of two ratings are plotted, there is perfect correlation if the points lie along any straight line. But there is perfect agreement only if the points lie along the line of equality (Bland & Altman, 1986). Since correlations are highly influenced by the range of values in the sample, it is essential to calculate the absolute reliability, including the standard error of measurement (SEM) and the smallest detectable difference (SDD). The precision of the measurement is expressed as the SEM. As postulated the SEMs of both raters were lower than 10 percent of the possible score range of the PIPS, in particular 2.6%, respectively 2.9%.

If the PIPS is used in intervention and longitudinal studies, the knowledge of the SDD is really important for early interventionists and researchers. The SDD value is important to determine whether the change scores rated by different raters for an individual child after training or maturation indicate real change (i.e., beyond measurement error) at the 95% confidence level. A statistically significant change between two ratings must be larger than the SDD (Portney & Watkins, 2009). The SDD for the PIPS in preschoolers with intellectual disabilities was 5.8 points or 7.2% of the possible score range of the PIPS, indicating that two assessments of the same child should differ by more than 6 points to reflect a factual imitation change. This small SDD value met Norman et al.'s (2003) criterion of an acceptable SDD. The SDD of 5.8 points was clearly below 0.5 standard deviation. This finding is consistent with the results of the SDD for the PIPS score in typically developing preschoolers (Vanvuchelen et al., in press-b).

Some implications of this study to clinical practise can be made. The SDD can be considered as a conservative estimate of patient's progress, identifying the smallest amount of change that could be interpreted as any improvement or decline. Therefore, using the SDD as a criterion for improvement may be thought of as having high specificity (avoiding false positives) but low sensitivity (finding many false negatives) (Portney & Watkins, 2009).

A critical remark has to be made. The present sample was comprised of Dutch speaking preschoolers with intellectual disabilities. However, two limitations should be noted. Directions for further research include: (1) the replication of the findings in other samples of preschoolers with intellectual disabilities, including those from other cultures and (2) the investigation of the utility of the PIPS in a randomised controlled trial to evaluate the effectiveness of different intervention programs for improving the imitation skills of young learners with imitation problems.

Acknowledgments

This study was partially supported by a Foundation Marguerite-Marie Delacroix grant to M. Vanvuchelen. We would also like to thank all children and the staff of the University Clinics of Antwerp and Ghent (Belgium) who participated in this research. We thank the staff of the home-based child development programs for children with mental disabilities and the staff of the special schools for children with learning disabilities, as well as the students Jolien Driessens (Department of Rehabilitation Sciences K.U.Leuven, Belgium) and Ellen Reynders (Department of Health Care, PHL University College, Belgium) for their assistance with data acquisition.

References

Atkinson, G., & Nevill, A. M. (1998). Statistical methods for assessing measurement error (reliability) in variables relevant to sports medicine. *Sports Medicine, 26*(4), 217–238.

Baer, D. M., Peterson, R. F., & Sherman, J. A. (1967). The development of imitation by reinforcing behavioral similarity to a model. *Journal Experimental Analysis of Behavior, 10*, 405–416.

Bland, J. M., & Altman, D. G. (1986). Statistical methods for assessing agreement between two methods of clinical measurement. Lancet, 8(1), 8476, 307–310.

Brown, A. K., Brown, J. L., & Poulson, C. L. (2008). Discriminating which fork to use: Teaching selective imitation to people with autism: Review. *Research in Autism Spectrum Disorders, 2*, 199–208.

Brown, F. J., Peace, N., & Parsons, R. (2009). Teaching children generalized imitation skills: A case report. *Journal of Intellectual Disabilities, 13*, 9–17.

Cardon, T. A., & Wilcox, M. J. (in press). Promoting Imitation in Young Children with Autism: A Comparison of Reciprocal Imitation Training and Video Modeling. Journal of Autism and Developmental Disorders. doi:10.1007/s10803-010-1086-8.

Fleiss, J. L. (1981). *Statistical methods for raters and proportions*. New York: Wiley.

Hurley, S., & Chater, N. (2005). Introduction: The Importance of Imitation. In S. Hurley & N. Chater (Eds.), *Perspectives on imitation. From neuroscience to social science. Volume 2: Imitation, human development, and culture* (pp. 1–52). Cambridge, MA: MIT Press.

Ingersoll, B. (2010). Brief report: Pilot randomized controlled trial of Reciprocal Imitation Training for teaching elicited and spontaneous imitation to children with autism. *Journal of Autism and Developmental Disorders, 40*, 1154–1160.

Ingersoll, B., & Gergans, S. (2007). The effect of a parent-implemented imitation intervention on spontaneous imitation skills in young children with autism. *Research in Developmental Disabilities, 28*, 163–175.

Ingersoll, B., & Lalonde, K. (2010). The impact of object and gesture imitation training on language use in children with autism spectrum disorder. *Journal of Speech, Language, and Hearing Research, 53*, 1040–1051.

Ingersoll, B., Lewis, E., & Kroman, E. (2007). Teaching the imitation and spontaneous use of descriptive gestures in young children with autism using a naturalistic behavioral intervention. *Journal of Autism and Developmental Disorders, 37*, 1446–1456.

Ingersoll, B., & Schreibman, L. (2006). Teaching reciprocal imitation skills to young children with autism using a naturalistic behavioral approach: Effects on language, pretend play, and joint attention. *Journal of Autism and Developmental Disorders, 36*, 487–505.

Kropmans, T. J, Dijkstra, P. U., Stegenga, B., Stewart, R., & de Bont, L. G. (1999). Smallest detectable difference in outcome variables related to painful restriction of the temporomandibular joint. *Journal of Dental Research, 78*, 784–789.

Lord, C., Rutter, M., DiLavore, P., & Risi, S. (2003). *Autism diagnostic observation schedule. Manual*. Los Angeles, CA: Western Psychological Services.

Lovaas, O. I. (1987). Behavioral treatment and normal educational and intellectual functioning in young autistic children. *Journal of Consulting and Clinical Psychology, 55*, 3–9.

Lyons, D. E., Young, A. G., & Keil, F. C. (2007). The hidden structure of overimitation. *Proceedings of the National Academy of Sciences of the United States of America, 104*, 19751–19756.

Macedoni-Luksic, M., Greiss-Hess, L., Rogers, S. J., Gosar, D., Lemons-Chitwood, K., & Hagerman, R. (2009). Imitation in fragile X syndrome. Implications for autism. *Autism, 13*, 599–611.

Masur, E. F. (2006). Vocal and action imitation by infants and toddlers during dyadic interactions: Development, causes and consequences. In S. J. Rogers & J. H. G. Williams (Eds.), *Imitation and the social mind. Autism and typical development* (pp. 27–47). New York: The Guilford Press.

Mehta, M., Pande, P., & Bhargava, M. (1991). Behavioral training for mothers of mentally handicapped children: Teaching of self-help skills. *Indian Pediatrics, 28*, 909–915.

Norman, G. R., Sloan, J. A., & Wyrwich, K. W. (2003). Interpretation of changes in health-related quality of life: The remarkable universality of half a standard deviation. *Medical Care, 41*, 582–592.

Perra, O., Williams, J. H. G., Whiten, A., Fraser, L., Benzie, H., & Perrett, D. I. (2008). Imitation and 'theory of mind' competencies in discrimination of autism from other neurodevelopmental disorders. *Research in Autism Spectrum Disorders, 2*, 456–468.

Portney, L. G., & Watkins, M. P. (2009). *Foundations of clinical research: Applications to practice* (3rd ed.). Upper Saddle River, NJ: Pearson/Prentice Hall.

Prinz, W. (2002). Experimental approaches to imitation. In A. N. Meltzoff & W. Prinz (Eds.), *The imitative mind: Development, evolution, and brain bases* (pp. 143–162). Cambridge: Cambridge University Press.

Rogers, S. J. (1998). Neuropsychology of autism in young children and its implications for early intervention. *Mental Retardation And Developmental Disabilities Research Reviews, 4*, 104–112.

Sevlever, M., & Gillis, J. M. (2010). An examination of the state of imitation research in children with autism: Issues of definition and methodology. *Research in Developmental Disabilities, 31*, 976–984.

Stahmer, A. C., Ingersoll, B., & Carter, C. (2003). Behavioral approaches to promoting play. *Autism, 7*, 401–413.

Tereshko, L., MacDonald, R., & Ahearn, W. L. (2010). Strategies for teaching children with autism to imitate response chains using video modeling. *Research in Autism Spectrum Disorders, 4*, 479–489.

Tomasello, M., Carpenter, M., Call, J., Behne, T., & Moll, H. (2005). Understanding and sharing intentions: The origins of cultural cognition. *Behavioral and Brain Sciences, 28*, 675–691.

Van Baalen, B., Odding, E., van Woensel, M. P., & Roebroeck, M. E. (2006). Reliability and sensitivity to change of measurement instruments used in a traumatic brain injury population. *Clinical Rehabilitation, 20*, 686–700.

Van der Meulen, B. F., Ruiter, S. A. J., Lutje Spelberg, H. C., & Smrkovsky, M. (2000). *Bayley scales of infant development. Nederlandse Versie. BSID-II-NL*. Lisse: Swets Test Publishers.

Vanvuchelen, M. (2009). *Imitation problems in children with autism spectrum disorders. A study of their nature, clinical significance and utility in diagnosis*. Ph.D. Dissertation in Rehabilitation Sciences and Physiotherapy, Group of Biomedical Sciences, Katholieke Universiteit Leuven, Belgium.

Vanvuchelen, M., Feys, H., & De Weerdt, W. (in press). Is the good-imitator-poor-talker profile syndrome-specific in Down syndrome?: Evidence from standardised imitation and language measures. *Research in Developmental Disabilities*. doi:10.1016/j.ridd.2010.09.010.

Vanvuchelen, M., Roeyers, H., & De Weerdt, W. (in press-a). Imitation Assessment and Its Utility to the Diagnosis of Autism: Evidence from Consecutive Clinical Preschool Referrals for Suspected Autism. *Journal of Autism and Developmental Disorders*. doi:10.1007/s10803-010-1074-z.

Vanvuchelen, M., Roeyers, H., & De Weerdt, W. (in press-b). Objectivity and Stability of the Preschool Imitation and Praxis Scale (PIPS). *American Journal of Occupational Therapy*.

Vanvuchelen, M., Roeyers, H., & De Weerdt, W. (2011a). Do imitation problems reflect a core characteristic in autism? Evidence from a literature review. *Research in Autism Spectrum Disorders, 5*, 89–95.

Vanvuchelen, M., Roeyers, H., & De Weerdt, W. (2011b). Development and initial validation of the Preschool Imitation and Praxis Scale (PIPS). *Research in Autism Spectrum Disorders, 5*, 463–473.

Williams, J. H., Whiten, A., & Singh, T. (2004). A systematic review of action imitation in autistic spectrum disorder. *Journal of Autism and Developmental Disorders, 34*, 285–299.

Zachor, D. A., Ilanit, T., & Itzchak, E. B. (2010). Autism severity and motor abilities correlates of imitation situations in children with autism spectrum disorders. *Research in Autism Spectrum Disorders, 4*, 438–443.